ChinaXiv合作期刊
计算机应用研究

chinaXiv:201804.01445v1

混合加噪声模型与条件独立性检测的因果方向推断算法 *

麦桂珍 [1a]，彭世国 [1a]，洪英汉 [1b, 2†]，陈平华 [2]，彭昱忠 [3]

(1. 广东工业大学 a. 自动化学院; b. 计算机学院，广州 510006; 2. 韩山师范学院 物理与电子工程学院，广东 潮州 521041; 3. 广西师范学院 科学计算与智能信息处理广西高校重点实验室，南宁 530001)

摘　要：从可观测的变量中推导出潜在的因果关系是人工智能领域的热点研究之一。传统的基于独立性检测的方法是通过检测 V 结构来确定一组马尔科夫等价类而非最终的因果关系；而加噪声模型算法却只能适应于低维度的因果网络结构。为此，提出一种采取分治策略的混合加噪声模型与条件独立性检测的因果方向推断方法。首先是将一个 n 维因果网络分解成 n 个诱导子网络，分别归入三种基本结构（单度结构、非三角结构和存在三角的结构）中的一种，从理论上分别证明其有效性；其次对每个诱导子网络进行基于加噪声模型算法与条件独立性检测相结合的方向推断；最后把所有子网络合并起来构建成完整的因果关系网络。实验表明，该方法比传统的因果关系推断方法更加有效。

关键词：因果网络；加噪声模型；马尔可夫等价类

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2017.12.0802

Causation inference based on combining additive noise model and conditional independence

Mai Guizhen[1a], Peng Shiguo[1a], Hong Yinghan[1b, 2†], Chen Pinghua[2], Peng Yuzhong[3]

(1*a. School of Automation*; *b. School of Computer Science & Technology Guangdong University of Technology*, *Guangzhou* 510006, *China*; 2. *School of Physice & Electronic Engineering*, *Hanshan Normal University*, *Chaozhou Guangdong* 521041, *China*; 3. *Key Laboratory of Scientific Computing & Intelligent Information Processing Guangxi Teachers Education University*, *Nanning* 530001, *China*)

Abstract: Inferring causal directions from observed variables is one of the fundamental problems in artificial intelligence (AI) field. Traditional conditional independence based methods usually learn causal directions by detecting V-structures and return Markov equivalence classes, instead of true causal structures; Most other direction learning methods can distinguish the equivalence classes, but are effective only in the bivariate (or two-dimensional) cases. This paper proposed a new approach for causal direction inference from general networks, based on a split-and-merge strategy. The method first decomposes an n-dimensional network into n induced subnetworks, each of which corresponds to a node in the network. Each induced subnetwork can be subsumed to one of the three substructures: one-degree, non-triangle and triangle-existence structures. Three effective algorithms are developed to infer causalities from the three substructures, and learning these induced subnetworks orderly to achieved the whole causal structure of the multi-dimensional network. Experiments show that the method is more general and effective than traditional methods.

Key words: causal networks; additive noise model; Markov equivalence classes

0　引言

从可观测的变量推断出事物中蕴含的因果关系是人工智能领域的研究热点之一。因果推断与主流的统计学习方法不同，它旨在研究数据的内在生成机制，而不是简单地研究观察变量的联合分布情况，如变量相关性。事实上，研究因果关系和相关性之间最重要的区别恰好就在于事物之间的因果先后顺序（即因果方向）的推断。事实证明，理解因果关系的方向对于预测任何可干预的结果和许多应用都是至关重要的，例如经济模型预测，基因分析。

收稿日期：2017-12-10；**修回日期：**2018-01-30　**基金项目：**国家自然科学基金资助项目（61374081，61562008）；广东省自然科学基金资助项目（S2013010013034）；广西自然科学基金资助项目（#GXNSFAA198228）；广东省科技项目（2014A030307049，2015A030401101，2015B090922014，2016B030306002，201604010099，2017A040405063，2016B030308001）

作者简介：麦桂珍（1985-），女，广东潮州人，博士研究生，主要研究方向为因果关系、机器学习；彭世国（1968-），男，教授，博士，主要研究方向为非线性分析、鲁棒控制、随机系统；洪英（1984-），男（通信作者），副教授，硕士，主要研究方向为因果关系、机器学习、云计算、数据挖掘及其应用（honyinghan@163.com）；陈平华（1967-），男，教授，硕士，主要研究方向为 Web 服务、Web 挖掘、云计算、推荐系统；彭昱忠（1980-），男，副教授，硕士，主要研究方向为智能计算、数据挖掘与建模.

ChinaXiv合作期刊
chinaXiv:201804.01445v1

从计算的角度来看，因果推断通常是采用一种概率图模型，模型中变量间的有向边表示因果方向[10]。由于对实验样本进行因果干扰是比较困难的，通常会采用条件独立（CI）测试来检测变量之间的局部因果关系[10,15]。通过条件独立测试来检测数据间蕴涵的 V 结构，就可以得到关于这些数据的一部分因果关系，这种方法是马尔可夫等价类方法[2]。尽管目前研究也利用一致性传播的算法[3]将给定组在马尔可夫等价类的基础上进行因果关系推断，但是数据不完整就会导致结果不准。以上方法能够找到的因果关系的准确性受制于从目标数据中检测到的 V 结构数量。如果给定数据不蕴涵 V 结构，这些方法则不能确定任何因果方向，仅能返回一个因果骨架，即无环图。

为了解决上述方法在因果推断中遇到的瓶颈，近十年来研究人员主要在下述这几种假设下并利用因果变量之间的不对称性对因果方向进行推断。其中有一种方法是 Shimizu 等人提出的线性非高斯加噪声模型（LINGAM）[13,14]，具体为：给定两个变量 x 和 y，首先假设它们之间的因果关系的方程为 $y=f(x)+\varepsilon$，其中 f 是一个线性函数，ε 是一个噪声项并统计独立于 x。其原理是在假定数据生成过程 f 是线性,且噪声 ε 是非高斯分布的情况下因果变量存在不对称性：a）ε 独立于 x；b）由于实际上还不知道 x 与 y 之间的因果方向，所以同时也得考虑相反的模型 $x=g(y)+e$，此时 e 不独立于 y。因此，通过对两个模型，自变量与噪声之间的独立性可以推断出 x 与 y 之间的因果方向。依据这一性质就可以对马尔可夫等价类进行因果方向推断。另一种方法是非线性加噪声模型(ANM)，适用于连续数据[6,18]和离散数据[11,12]。类似地，ANM 的原理如下：给定两个变量 x 和 y，用来表达他们之间的因果关系的方程 $y=f(x)+\varepsilon$，其中 $f(*)$ 是一个非线性函数，ε 是一个噪声项并统计独立于 x。其最核心的理论在于 $P(x,y)$ 在绝大部分情况下只允许 $x\rightarrow y$，也就是正方向的 ANM，当考虑反方向的 ANM，$x=g(y)+e$，会发现残差项 $e'=x-E(x\mid y)$ 与自变量 y 是不独立的，因此 $x\leftarrow y$ 不成立。总的来说，基于加噪声模型是通过检测联合分布 $P(x,y)$ 在正反 ANM 间的不对称性来推断出因果关系。PostNonLinear(PNL)模型[15]进一步泛化了 ANM 在数据生成过程中的假设,它允许一个额外的双射转换数据，具体地说,该模型具有形式 $y=g(f(x))+e$，其中 $g:R\rightarrow R$ 是双射。

近年来，一些因果方向推断方法也通过信息几何模型来挖掘因果变量之间的不对称性[4][7][8]。这些方法主要是：在因果方程 $y=f(x)+\varepsilon$ 中，在自变量 x 与生成机制 $f(x)$ 是独立的假设下进行因果方向的推断。虽然这些方法能以不同的方式检测数据之间不对称的原因和影响,但它们仅在二维的情况下有效。类似的，ANM 和 PNL 在绝大部分情况下也仅在二维有效[16,18]。另外，现有的一些混合模型算法，如 CDHD[5]，在某种程度上可以从多维网络中发现因果关系。然而有研究表明，CDHD 方法并不十分准确，因为它直接使用了一个双变量因果学习方法 IGCI[7,8]，在方向学习阶段直接对目标节点的诱导子图进行方向推断，从理论上这是一种错误的做法，因为如果诱导子图的维度总大于 3，理论上此方法的准确率就会趋向于 50%，也就是等同于随机判别。

综上所述，传统的基于 V 结构的方法、加噪声模型、信息几何模型等都无法有效地从高维观测数据中推断出因果关系。本文提出了一种采取分治策略的混合加噪声模型与条件独立性检测的因果方向推断方法。该方法首先是将一个 n 维因果网络分解成 n 个诱导子网络，每个子网络对应于网络中的一个节点。其中诱导子网络都可以被归入三个基本结构（单度结构、非三角结构和存在三角的结构）中的一个，理论上证明了三种基本结构的有效性；其次对每个诱导子网络进行基于加噪声模型算法与条件独立性检测相结合的方向推断；最后把所有子网络合并起来构建成完整的因果关系网络。实验表明，本文提出的方法在因果关系推断中比传统的方法更加有效，适应性更强。

本文的贡献如下：a)在含有三角结构的因果网络中，提出了一种新的因果关系推断方法可以大大缩小时间复杂度的算法；b)将因果网络分割成三个基本的子网络进行因果关系推断。并且进行了大量实验来评估该方法，实验结果表明，本文提出的方法（VE+ANM）性能优于现有的经典方法。

1　准备知识

考虑一个可以表示为无向图 $G_N=\{V_N,E_N\}$ 的网络，其中 $V_N=\{x_1,x_2,\cdots,x_n\}$ 和 $E_N=\{e_1,e_2,\cdots,e_n\}$ 表示网络中的节点和边缘。对于任何节点 $y\in V_N(y=x_k)$，方向学习的目标是发现 y 和邻点之间的因果关系的方向集合。所有的边都是面向的，得到一个有向无环图形(DAG) $G_D=\{V_D,E_D\}$ 的位置在 E_D 中，$V_D=V_N$ 和每一个有向边对应于 E_N 的一个唯一的无向边。本文称 G_D 为 G_N 的推导网络。对于任何 N 中的节点 y，可以得到一个诱导子图 $G_y=\{V_y,E_y\}$，其中 V_y 由 y 和它的邻居集合 N_y 组成。V_y 可以根据下面的方式分为两个子集 NT_y 和 T_y，对于任意节点 $x\in V_y$，如果 x 的度大于 1，则将 x 移到 T_y，否则将 x 移到 NT_y。这样本文可以将 V_y 纳入以下三个子结构中的一个：

定义 1　如果 $|N_y|=1$，则称 G_y 是单度结构，简称 ODS。

定义 2　如果 $|N_y|>1$ 而且 $|T_y|=0$，则称 G_y 是非三角结构，简称 NTS。

定义 3　如果 $|N_y|>1$ 而且 $|T_y|>0$，则称 G_y 是存在三角的结构，简称 TES。

定义 4　V 结构和三角结构。对于 G_D 中的三个节点 x，y 和 z，如果 x 和 z 是 y 的父亲节点，而且 x 和 z 之间不存在边，则可以说 x、y、z 构成一个 V 结构；反之，如果 x 和 z 之间存在一条边，则说 x、y、z 构成一个三角结构。

定义 5　若 G_y 中存在 V 结构，则称 G_y 为 VES。

定义 6　若 G_y 中不存在 V 结构，则称 G_y 为 NVS。

ChinaXiv合作期刊

定义 7 三个变量（或节点）x、y 和 z，如果 x 和 y 满足结构方程 $y=f(x)+\varepsilon$，其中 ε 是一个噪声项，可称 x 和 y 构成一个确定性的情况。如果 $y=f(x,z)+\varepsilon$，则称 x 和 y 构成一个不确定性的情况。

定义 8 D 分离。设 P 是变量 x_i 到变量 x_j 的路径，P 被一组变量 $Z(Z\subseteq V_D)$ 阻断，当且仅当符合以下其中一个情况：a）P 包含一个链，$x_i\leftarrow x_k\leftarrow x_j$，且 $x_k\in Z$；b）P 包含一个分叉，$x_i\leftarrow x_k\rightarrow x_j$，且 $x_k\in Z$；c)P 包含一个对撞，$x_i\rightarrow x_k\leftarrow x_j$，$x_k\notin Z$ 且 x_k 没有任何后裔存在于 Z 中。

2 方法

2.1 V 结构搜索算法

给定一个无向图 $G_N=\{V_N,E_N\}$，其中包含一个诱导子图 $G_y:G_y=\{V_y,E_y\}$（V_y 由 y 和它的邻居节点集合 N_y 组成)。如果 G_y 属于 NTS，则 $\forall x\in N_y$ 只与 y 直接相连。

本文提出的方法采用 V 结构搜索和 ANM 相结合的方法，发现 y 与 N_y 之间的方向。注意到，大多数基于 V 结构的方法在 CI 测试中具有指数复杂度，在考虑小图形时，很难避免产生大量冗余的计算。还有一些 V 型结构的识别方法更有效率，如 SVS [1]，它根据下面的定理来确定 V 型结构：

定理 1 给定三个变量 x_1、x_2 和 y，$x_1\rightarrow y\leftarrow x_2$ 是一个 V 结构，当且仅当存在变量集 $Z\subseteq G_N\setminus(x_1,x_2)$ 使得 1）$x_1\perp x_2|Z$；2）$x_1\not\perp x_2|(y,Z)$。

上面的定理说明了如何普遍地找到 V 结构。然而很难从数百个变量中找到一个可能的条件集合 Z (或更多，如高维网络)。因此，这些方法用在本文提出的方法中是无效的或不适用的。在节中首先提出了一种有效的方法，从 $G_y(NTS)$ 中寻找 V 型结构，如以下推论所示。

推论 1 给定一个有向无环图 $G_D=\{V_D,E_D\}$，$x,y,z\in V_D$，如果 x,y,z 形成一个 V 结构 $x\rightarrow y\leftarrow z$，则 $\exists N_{xz}$，$N_{xz}\subset(N_x\cup N_z)$，使得 $x\perp z|N_{xz}$ 和 $x\not\perp z|(N_{xz}\cup y)$。

证明 1）如果 x 和 z 都是根。令 $N_{xz}=\varnothing$，则有 $x\perp z$ 和 $x\not\perp z|y$。2）如果 x 或 z 是多度的，则在 x 和 z 之间可能存在三种路径：1，$x\leftarrow\cdots\leftarrow z$（或者 $x\rightarrow\cdots\rightarrow z$）；2，$x\leftarrow\cdots\rightarrow z$；3，$x\rightarrow\cdots\leftarrow z$。令 a 为 x（$a\in N_x$）的一个邻居节点，一条途径 p 覆盖 $\{x,a,z\}$，如果 P 属于第一种路径或者第二种路径，必须有 p 被 a 挡住；如果 p 属于第三种路径，则一定有一个对撞 b，且 p 也被 $\varnothing$ 或者 a（$a\in Nx$）挡住，a 不是 b 的后代。总的来说，本文可以找到这样的一组 a（记为 N_{xz}，$N_{xz}\subseteq N_x\cup N_z$）来阻止 x 到 z 或者 z 到 x 的任何路径，即 $x\perp z|N_{xz}$。又由于 y 是 x 和 z 之间的一个对撞，则对于任意的 D 分离的集合 N_{xz} 有 $y\notin N_{xz}$。因此,x 和 z 不能被 $N_{xz}\cup y$ D 分离。

根据上面提到的,当考虑结构 $x\rightarrow y\rightarrow z$ 或 $x\leftarrow y\leftarrow z$ 时很容易获得以下推论，如下：

推论 2 给定一个有向无环图 $G_D=\{V_D,E_D\}$，$x,y,z\in V_D$，如果 x,y,z 形成一个结构 $x\rightarrow y\rightarrow z$ 或 $x\leftarrow y\leftarrow z$，则 $\exists N_{xz}$，$N_{xz}\subset(N_x\cup N_z)$，使得 $x\perp z|N_{xz}$。

推论 3 给定一个无向图 $G_N=\{V_N,E_N\}$,一个结构 $x-y-z$（$x,y,z\in V_N$,x 到 z 不是直接相连的)；如果 $\exists N_{xz}\subseteq(N_x\cup N_z)$，使得 $x\perp z|N_{xz}$ 和 $x\not\perp z|(N_{xz}\cup y)$（$y\notin N_{xz}$），则那么 $x-y-z$ 在联合密度 $P(V_N)$ 下对应于一个 V 结构。

证明 根据推论 1 和 2，$\exists N_{xz}\subseteq(N_x\cup N_z)$ 且 $x\perp z|N_{xz}$，如果 $x-y-z$ 并不对应于联合密度 $P(V_N)$ 下的 V 结构，即 $x\rightarrow y\rightarrow z$ 或 $x\leftarrow y\leftarrow z$，则有 $y\notin N_{xz}$ 满足 D 分离条件（阻断路径 $x-y-z$）。即 y 必须包含在 D 分离集合 N_{xz} 中，这与 $y\notin N_{xz}$ 矛盾。因此 $x-y-z$ 只能构成一个 V 结构。

可以看到，推论 1 和 3 说明了如何在 $\{x,y,z,N_x,N_y\}$ 中发现一个 V 结构，并可以归纳为如下推论 4。

推论 4 给定一个无向图 $G_N=\{V_N,E_N\}$，且包含子图 $G_y=\{V_y,E_y\}$，一个结构 $x-y-z$ 在联合密度 $P(V_N)$ 下对应于一个 V 结构，当且仅当满足以下两个条件：a）$\exists N_{xz}$, $N_{xz}\subset(N_x\cup N_z)$，使得 $x\perp z|N_{xz}$；b）$\exists N_{xz}$, $N_{xz}\subset(N_x\cup N_z)$，使得 $x\not\perp z|(N_{xz}\cup y)$。

可以看到推论 4 显著缩小了定理 1 中描述的条件集 Z 的搜索范围。因此，能更高效地识别 V 型结构，尤其是在高维网络中。而 V 结构通常广泛存在于高维网络中,通过使用推论 4，可以检验在目标 G_y 中给定的一个结构 $x-y-z$ (x 和 z 不是直接连接) 是否满足推论中的两个条件,如果是就可以推断出 $x\rightarrow y\leftarrow z$。

然而仍然有一个问题，底层的 CI 测试方法在最坏的情况下必须运行 $2^{|N_x\cup N_z|}$ 次。直观上，两个节点通常不会被大量的共同邻居节点 D 分离，从计算方面来说，当条件集足够大时，CI 测试方法很难保障其准确率。因此，在 CI 测试中，本文设置了一个阈值 k 来限制 N_{xz} 的数量，可以根据先验信息或实验结果通过 CI 测试方法和数据分布类型来选择 k。对 V 型结构的搜索算法总结在算法 1 中。

算法 1

Algorithm 1 Searching V-structure

Input: $G_y=\{V_y,E_y\}$, threshold k

Output: V-structures

1: **for** $\forall$ node pair x, z ($x,z\in N_y$, x and z are not directly connected) **do**

2: Remove $\forall node\in(N_x,N_z)$ beside the adjacency paths between x and z.

3: **for** $\forall N_{xz}\subset(N_x\cup N_z)$, $|N_{xz}|<k$ **do**

4: **if** $x\perp z|N_{xz}$ and $x\not\perp z|(N_{xz}\cup y)$ **then**

5: let x, z be the parents of y (V-structures)

6: **end if**

7: **end for**

chinaXiv:201804.01445v1

```
end for
```

2.2 NTS、ODS、TES 学习算法

考虑 NTS 结构的实现过程：

运用算法 1 在目标节点 y 及其邻居节点集 N_y 之间进行搜索，例如如果算法 1 判断 $x-y-z$ 为一个 V 结构，则判断 x，z 为 y 的父亲节点。根据定义 2，NTS 结构中不包含三角结构，通过算法 1 可以找到 NTS 中所有的 V 结构，从而关于 y 的所有父亲节点可以得到识别。类似的方法在很多其他的文献中已经有很广泛的运用[19]。

考虑 ODS 结构的实现过程：

根据定义 1，ODS 结构只包含两个节点，因为叶子节点与其父亲必然构成确定性关系，也就是 $x=f(y)+e$，它满足 ANM 条件 $y\to x$。因此，利用 ANM 中回归残差与自变量独立的特性，也就是 e 独立于 y，可以很容易推断出目标节点的儿子为 x；反之，则 x 为 y 的父亲节点。

考虑 TES 结构的实现过程：

TES 是因果方向推断中的难题，在以往的文献中[20],需要遍历所有节点找出符合 ANM 的 y 及其一组父亲节点，也就是 $y=f(PA_y)+e$，其中 e 独立于节点集 PA_y 中的每一个变量。这种方法有两大问题：一是满足 ANM 的父亲节点集有可能不是唯一的，找到的 PA_y 仅仅是 y 的一部分父亲节点；二是枚举法在$|N_y|$较大时，计算机无法在可接受的时间内返回结果。本文的工作要克服的最重要一个难题就是如何在 TES 中快速寻找 y 的所有父亲节点。本文提出定理 2：

定理 2 若 G_y 至少存在一个 V 结构，即 VES 结构，则 1）y 的所有父亲节点均包含在 V 结构中；或 2）如果存在 y 的一个父亲节点 Pa_y1 不包含在任何 V 结构中，比如 V 结构 $Pa_y2\to y\leftarrow Pa_y3$，若 Pa_y1 包含在 Pa_y2，Pa_y3 的条件独立集中，则边方向为 $Pa_y1\to y$。

证明 根据定义 1、2、3,知道 G_y 必为 ODS、NTS、TES 中的一个。如果存在 V 结构，则 G_y 只能是 NTS 或者 TES。若 G_y 属于 NTS，则根据 V 结构定义 4，可以直接推断出 y 的所有父亲节点均包含在 V 结构中。若 G_y 属于 TES，用反证法，假设存在 y 的一个父亲节点 Pa_y1 不包含在关于 y 的 V 结构中，由于 G_y 最少包含一个 V 结构，假设其为 $Pa_y2\to y\leftarrow Pa_y3$，则 Pa_y1 与 Pa_y2 必然存在一条直接相连的边，否则 $Pa_y1\to y\leftarrow Pa_y2$ 形成了一个 V 结构，矛盾。同理可以知道 Pa_y1 与 Pa_y3 同时存在一条直接相连的边。由于 $Pa_y2\to y\leftarrow Pa_y3$ 是一个 V 结构，因此 Pa_y1、Pa_y2 之间不存在直接相连的边，若 Pa_y1 包含在 Pa_y2、Pa_y3 的条件独立集中，则若 Pa_y1、Pa_y2、Pa_y3 三者关系为：$Pa_y2\to Pa_y1\to Pa_y3$，$Pa_y2\leftarrow Pa_y1\to Pa_y3$ 或 $Pa_y2\leftarrow Pa_y1\leftarrow Pa_y3$。根据因果网络推断的一致性传播特性，可以知道 Pa_y1 与 y 的因果关系为 $Pa_y1\to y$，否则存在局部环结构。

定理 2 指出，如果一个 TES 存在 V 结构，也就是一个 VES，则在大部分情况下，关于 y 的父亲节点都可以根据条件独立性检测得到的。因为根据定理 2 的证明，仅在 $Pa_y2\to Pa_y1\leftarrow Pa_y3$ 的情况，Pa_y1 与 y 的因果关系无法区分，其余所有情况，TES 里的因果关系都是可以通过 V 结构蕴涵的条件独立性得到识别的。另一方面，如果 TES 不存在 V 结构，则这种情况只能是根据 ANM 进行识别。在大量数据集中，完全不包含 V 结构的 TES 在真实网络中是很少见的，也就是说在一般的真实网络中，TES 都可以得到有效的识别。根据上述对 NTS、ODS、TES 三个结构的讨论，本文给出一个有效的对高维因果网络进行方向学习的算法。

a)将 n 维网络 G_n 分解为 n 个关于每一个节点的诱导子图 G_y，将 G_y 集分类，分别为 NTS，TES 和 ODS。

b)在 NTS 中，根据定理 2 首先在 N_y 中搜索关于 y 的所有 V 结构；然后考虑剩余非 V 结构节点 x_k，如果 x_k 存在于任意一个 V 结构（不包含 y）的条件独立集中，则将 x_k 归类到 y 的父亲节点集；最后利用 ANM 找出关于 y 的其他可能的父亲节点。

c)在 TES 中，根据定义 3，判断 x_i-y-x_j 是否为一个 V 结构，则判断 x_i，x_j 为 y 的父亲节点。

d)在 ODS 中，根据 ANM 的性质，因为 ODS 结构只包含两个节点，叶子节点与其父亲必然构成确定性关系，也就是 $x=f(y)+e$，满足 ANM 条件 $y\to x$。因此可利用 ANM 中回归残差与自变量独立的特性，也就是 e 独立于 y，可以很容易推断出目标节点的儿子为 x；反之，则 x 为 y 的父亲节点。

e)将所有判明方向（关于 y 父亲节点已经被识别的）的诱导子图合并，得到一个完整的有向无环图 G_D。

2.3 提出方法（VE+ANM）的框架

算法 2 （VE+ANM）框架

Algorithm 2 The framework of (VE+ANM)

Input: undirected graph $G_N=\{V_N, E_N\}$

Output: DAG $G_D=\{V_D, E_D\}$

```
Divide G_N into |V_N| induced subgraphs, one of which corresponds to one and only one of the substructures ODS, NTS or TES
for each induced subgraph G_y = {V_y, E_y} do
  if G_y : TES then
    for ∀x_i, x_j ∈ N_y do
    search V-structure: x_i → y ← x_j
    if such x_i → y ← x_j does exist then
  let x_i and x_j be PA_y , others be CH_y ;
      else if x_k in N_y is contained in the separated set of x_i and x_j
        Let x_k be the PA_y
else
     employs ANM to find PA_y ;
     if PA_y ≠ ∅ then
       let N_y \ PA_y be CH_y ;
```

chinaXiv:201804.01445v1

ChinaXiv合作期刊
chinaXiv:201804.01445v1

```
12:        else
13:            let N_y \ T_y be CH_y ;
14:        end if
15:      end if
16:    end for
17: end if
18: if G_y : NTS then
19:   the same to line 4-16 (T_y = ∅) .
20: end if
21: if G_y : ODS then
22:    employs ANM to infer the direction.
23:    end if
24: end for
25: Merging all the partial results (plus a refinement
step if exists any un-oriented edge), out put G_D
```

3 实验

在本章实验中将采用由 12 个真实的因果网络生成的数据集来对本文提出的方法进行评估。这些真实网络涉及各个领域，包括生物与医学(Cancer、Asia, Spiegel-HalterDLC93、Alarm、Diabetes、Mildew)、车辆诊断(Car-diagnosis)、石油勘探(Oil-wildcatter)、场景分析(Boerlage92)、天气预报(Hailfinder)、打印机故障排除(Win95pts)和基因网络(Link)，数据集由这些真实网络基于因果推断领域的经典方法[12]生成的。

在本节实验中，将(VE+ANM)方法与 4 种主流的因果推断算法进行比较，其中包括基于 V 结构的因果推断算法 SVS、基于 V 结构与一致性传播结合的因果推断算法 POE、加噪声模型因果推断算法 ANM 和信息几何模型因果推断算法 IGCI。本文以实验报告的准确率为评价标准，并定义准确率为：

$$\text{准确率}=\frac{\text{推断的方向}\cap\text{实际的方向}}{\text{实际的方向}}。$$

3.1 与 SVS、POE、ANM 和 IGCI 比较：

如表 1 所示，在 12 个网络中(VE+ANM)方法的准确率明显优于其他对比方法。(VE+ANM)方法在小网络(如 Asia)上的准确性已经达到了 94%，在大网络的准确率也达到了 80%和 78%，这表明本文提出的方法在不同维度网络都具有良好的鲁棒性。特别地，注意到(VE+ANM)方法在网络 Link 的性能是最差的(只有 78%)。这是因为该网络有 40%的边包含在三角形结构中，这使得该网络的结构比其他网络结构更加复杂。注意到，在本实验中样本量都达到 2 000，大部分算法都接近了最理想的实验环境，因此这些准确率代表的是这些方法在准确率上的峰值。

另一方面，可以看到网络从低维到高维，除开 ANM 会受到维度影响外，其余算法都相对稳定，准确率并没有受到维度的影响。特别地，IGCI 在多节点情况下趋向于随机判断，因此准确率几乎就在 60%左右波动。而 SVS 与 POE 是基于 V 结构识别的算法，因此理论上不会受到维度的影响。可见，尽管(VE+ANM)方法运用了残差与自变量独立的原理进行识别，但根据本文的理论，此方法也不会被高维影响，在所有网络结构下都保持了较高的准确率。

可以看到，无论 SVS 还是 POE 都不能相对准确地发现对应的实际图形，特别是当网络结构很小且只有很少 V 结构的时候(如 Cancer、 Asia)，因为总有一些不定向的边未不能被一致性地扩展。以 Asia 网络为例(图 1)，两个 V 结构 $x_3 \to x_6 \leftarrow x_4$ 和 $x_6 \to x_8 \leftarrow x_5$ 是可以被推断出来的,那么使用一致的扩展可以很容易获得 $x_6 \to x_7$，然而其他三条边不能通过任何一致性的扩展来定向。而 SVS 和 POE 之间的区别仅仅是识别 V 结构的方法，它们的性能很大程度上受到 V 结构数量的影响，由于(VE+ANM)方法有更强的适应性所以在多维网络上有更好的表现。

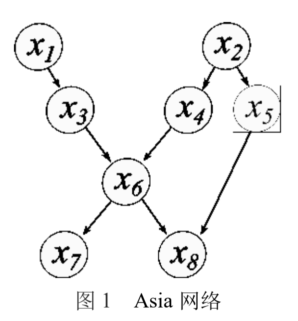


图 1 Asia 网络

ANM 表现得与其他方法截然不同。如表 1 所示，可以看出 ANM 的准确率受复合确定性关系的节点对数影响，其主要原因是 ANM 在许多不确定的情况下无法得出关于方向的结论，也无法确定这些边的方向。然而这种情况却广泛存在于多维网络中。类似地，IGCI 只适用于双变量情况，它的性能也取决于节点对数。特别地，IGCI 推断的结果有时会比随机选择差(如 SHDLC93、Boerlage92)。这是因为信息几何模型不适用于多维（$n>2$）网络。

表 1 在 12 个网络的运行结果

网络	准确率				
	(VE+ANM)	SVS	POE	ANM	IGCI
Cancer	0.89	0.55	0.57	0.66	0.66
Asia	0.94	0.74	0.79	0.61	0.64
Oil-wildcatter	0.85	0.82	0.82	0.48	0.69
Car-diagnosis	0.92	0.84	0.81	0.42	0.71
SHDLC93	0.84	0.82	0.83	0.48	0.49
Boerlage92	0.83	0.82	0.82	0.45	0.46
Mildew	0.88	0.83	0.87	0.12	0.63
Alarm	0.87	0.85	0.83	0.33	0.61
Hailfinder	0.91	0.81	0.84	0.44	0.61
Win95pts	0.92	0.85	0.82	0.25	0.70
Diabetes	0.80	0.75	N.A.	0.26	0.52
Link	0.78	0.77	N.A.	0.29	0.61

chinaXiv:201804.01445v1

3.2　在三种子结构上的效果比较

在本节实验中采用两个结构相对复杂的真实网络结构，Hailfinder 网络 (其中有 12%的边包含在三角形结构中)和 Oil-wildcatter 网络(其中有 48%的边包含在三角形结构中)以及{30,50,100,200,500,1000,5000}样本下处理三种子结构 ODS、NTS、TES 来评估(VE+ANM)方法。

图 2 所示是在 Hailfinder 网络下对(VE+ANM)进行评估。其中 ODS 的性能略优于 NTS 和 TES，当样本容量超过 1 000 时，其准确率能达到 96%。这是因为样本尺寸相对较小时，可能会根据 ODS 的方向来更新 ODS 与 TES 或 ODS 与 NTS 之间的冲突。在 ODS 案例中，本文只测试两个变量之间的依赖关系，不需要任何条件集，且 ODS 的结构非常简单。虽然 NTS 和 TES 的性能受到 CI 测试方法的显著影响，但 CI 测试的准确性依赖于给定的(局部)结构。因此，当结构相当复杂时，CI 测试总是需要较大的样本量。可以看到，当样本量达到很大的时候 ODS、NTS 和 TES 的曲线趋向于重叠。此外，NTS 比 ODS 和 TES 需要更多的样本，可以看到样本大小从 30 到 5 000 个，NTS 的准确性的变化最明显，从 52%到 94%。

图 3 所示是在另一个网络 Oil-wildcatter 评估(VE+ANM)方法。可以看到四个曲线的上升趋势与在 Hailfinder 网络的性能相似。但是 TES 的性能比其他的要差。特别注意到，在该网络下有 40%的边在三角形结构中，且存在一些边不能定向或做任何一致性扩展。

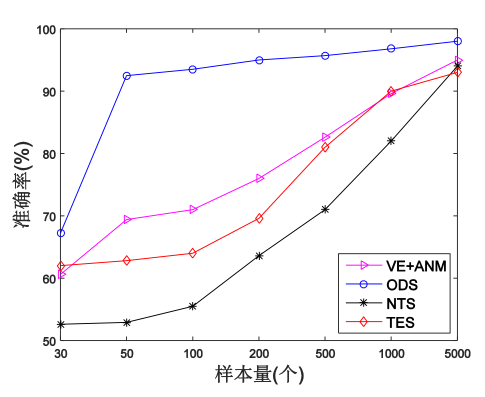


图 2　在 Hailfinder 网络各方法的性能

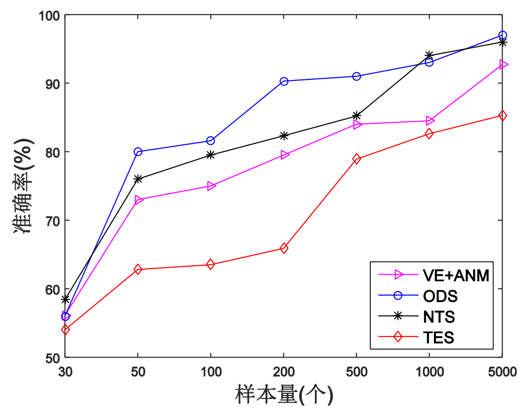


图 3　在 Oil-wildcatter 网络各方法的性能

实验表明，VE+ANM 混合方法比现有的单一方法明显有更高的鲁棒性，更能适用于复杂网络的因果推断问题上。然后受限于 TES 的高复杂性，VE+ANM 在 TES 上面的表现要远低于另外两种结构，一方面是多次条件独立性测试容易产生连锁的错误传递效应，另一方面是网络结构一旦复杂，很可能有一些子结构会违反因果忠实性与马尔可夫性，导致条件独立性不等价于 D 分离性，这目前仍然是因果推断领域的难题。

4　结束语

本文提出了一种通用的可扩展的方法，即(VE+ANM)方法，使用分裂合并策略，支持多维网络的因果方向推断。虽然大多数现有的方法都是建立在 V 型结构的学习或不对称检测上，但(VE+ANM)方法考虑到三种可能的子结构类型，从一个任意的网络中发现了一个更一般的因果方向。较强的理论分析证明了在一般多维网络下(VE+ANM)方法的有效性和准确性。虽然有时(VE+ANM)在某些特殊情况下也无法达到完整性，但实验结果验证了本文方法比最先进的方法更为普遍和有效。

参考文献：

[1] Cai R, Zhang Z, Hao Z. Causal gene identification using combinatorial v-structure search [J]. Neural Networks, 2013, 43: 63-71.

[2] Cheng J, Bell D, Liu W. Learning Bayesian networks from data: An efficient approach based on information theory [J/OL]. (1998) . http://www. cs. ualberta. ca/~ jcheng/bnpc. htm.

[3] Chickering D M. Learning equivalence classes of bayesian-network structures [J]. The Journal of Machine Learning Research, 2002, 2: 445-498.

[4] Daniusis P, Janzing D, Mooij J, *et al.* Inferring deterministic causal relations [J]. Machine Learning, 2012, arXiv: 1203. 3475.

[5] Hao Z, Zhang H, Cai R, *et al.* Causal discovery on high dimensional data [J]. Applied Intelligence, 2014, 42 (3): 594–607.

[6] Hoyer P O, Janzing D, Mooij J M, *et al.* Nonlinear causal discovery with additive noise models [C]// Proc of the 21st International Conference on Neural Information Processing Systems. 2009: 689–696.

[7] Janzing D, Mooij J, Zhang K, *et al.* Information-geometric approach to inferring causal directions [J]. Artificial Intelligence, 2012, 182: 1–31.

[8] Janzing D, Steudel B, Shajarisales N, *et al.* Justifying information-geometric causal inference [M]. [S. l.] : Springer International Publishing, 2014: 253–265.

[9] Mooij J, Janzing D, Peters J, *et al.* Regression by dependence minimization and its application to causal inference in additive noise models [C]// Proc of the 26th Annual International Conference on Machine Learning. 2009: 745–752.

[10] Pearl J. Causality: models, reasoning, and inference [M]. [S. l.] : Cambridge University Press, 2000: 639.

[11] Peters J, Janzing D, Sch¨olkopf B. Identifying cause and effect on discrete data using additive noise models [C]// Proc of International Conference on

ChinaXiv合作期刊

Artificial Intelligence and Statistics. 2010: 597-604.

[12] Cai R, Zhang Z, Hao Z. Sada: a general framework to support robust causation discovery [C]// Proc of the 30th International Conference on International Conference on Machine Learning. 2013: 208–216.

[13] Shimizu S, Hoyer P O, Hyv¨arinen A, *et al.* A linear non-gaussian acyclic model for causal discovery [J]. Journal of Machine Learning Research, 2006, 7 (4): 2003–2030.

[14] Shimizu S, Inazumi T, Sogawa Y, *et al.* Directlingam: a direct method for learning a linear nongaussian structural equation model [J]. Journal of Machine Learning Research, 2011, 12 (2): 1225-1248

[15] Spirtes P, Glymour C N, Scheines R. Causation, prediction, and search [M]. [S. l.] : Springer New York, 1993: 272-273.

[16] Zhang K, Hyv¨arinen A. On the identifiability of the post-nonlinear causal model [C]// Proc of the 25th Conference on Uncertainty in Artificial Intelligence. [S. l.] : AUAI Press. 2009: 647–655.

[17] Zhang K, Peters J, Janzing D, *et al.* Kernel-based conditional independence test and application in causal discovery [J]. Computer Science. 2012, 6 (8): 895-907.

[18] Peters J, Mooij J M, Janzing D, *et al.* Causal discovery with continuous additive noise models [J]. Journal of Machine Learning Research, 2013, 15 (1): 2009-2053.

[19] 张浩，郝志峰，蔡瑞初，等. 一种适用于高维网络的方向推断算法 [J]. 小型微型计算机系统, 2015, 36 (6): 1358-1362.

[20] Mai Guizhen, Hong Yinghan, Peng Shiguo, *et al.* Inferring causal direction from multi-dimensional causal networks for assessing harmful factors in security analysis [J]. IEEE Access. 2017, 5: 20009-20019.

[21] Zhang K, Hyv¨arinen A. Distinguishing causes from effects using nonlinear acyclic causal models [J]. Journal of Machine Learning Research, 2008, 6: 157–164.

[22] Peters J, Janzing D, Sch¨olkopf B. Causal inference on discrete data using additive noise models [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2011, 33 (12): 2436–2450.

chinaXiv:201804.01445v1